White paper
Public

Towards a Multimodal Transportation Data Framework

Supporting emerging transportation models and services

Contents

Executive summary

Digitisation is occurring at a rapid pace across all modes of transportation. Whether this be future disruptive services such as Mobility-as-a-Service or improving road safety through Connected and Automated Vehicles; it has become apparent that the secure exchange of data is critical to the future success of transportation.

In June 2018, a number of workshops were held across Australia with over 40 Transportation stakeholders representing agencies, operators, users and vendors. These groups provided insight into the role and relationship of data within the industry and this whitepaper represents the collective thoughts of the team captured during the workshops.

The Multimodal Transportation Data Framework represents a construct that enables four key capabilities required in order to support the data needs of the industry: Data Exchange; Data Brokerage; Data Catalog and Digital Rights and an over-arching Governance structure supporting a broad number of producers and consumers within a community geographically dispersed across the country and abroad. It represents the capture and support of different data requirements such as real-time or scheduled data and the need to evolving service requirements.

Rapid advancement within the industry is driving the further need to deploy a framework and trial the system immediately. Multiple use cases already exist such as Connected and Automated Vehicles (CAV), Intelligent Congestion Management, Mobility-as-a-Service (MaaS) and Condition Based Monitoring.

The recommendations include leveraging similar distributed and large scale Data Exchange infrastructure such as Internet Exchange Points (IXP's) and the role that government plays in enabling the foundation construct. Moreover, it is imperative that trials begin immediately to support the emerging transport services and leverage the digital transformation that is disrupting the industry today.

Introduction

Transportation modes are becoming increasingly integrated, connected, and in the near future, automated. There is the potential to explore new opportunities, combining data and insights from multiple sources to lead towards more intelligent, efficient use of multimodal transportation assets and creating new business models.

There are many instances around the world, where transport assets have been developed with a singular modality or a specific service as the primary focus. For example, modern high speed rail without trunked feeder systems, automated ports with legacy roads and rail, mass transit systems without connection protection or last mile solutions.

In the public transport sector, there is significant interest in moving some bus services from a scheduled timetable to on-demand: one that adapts and reacts based on customer travel patterns. This may see additional bus services being provided not just on a time-of-day/day-of-week basis but also in reaction to events such as a problems with other modes of transport (a train that has broken down). Improving 'first-mile/last-mile' access, with coordination between public and private transportation solution providers and rail services is another goal expressed by many authorities.

Taking a 'multimodal' approach requires thinking and planning for a future where being able to exchange data safely and securely, will begin to create actionable intelligence, providing a safer, frictionless, experience for the traveller, shipper or buyer. To this end we want to consider the city, state, region, or country as a 'platform' to deliver this experience.

Mobility-as-a-Service, 'describes a shift away from personally-owned modes of transportation and towards mobility solutions that are consumed as a service. This is enabled by combining transportation services from public and private transportation providers through a unified gateway that creates and manages the trip'[1]. In order to realise this concept, the exchange of data is fundamental.

Today, data is exchanged between certain parties such as transport authorities and regulated transportation providers. This exchange typically occurs within a closed, controlled environment. Some data may be considered to be 'open' and is subsequently made available to the public by the transport authority but there are few examples of facilities that enable the exchange of data between public and commercial entities.

[1] https://en.wikipedia.org/wiki/Transportation_as_a_Service

There are many sources of data in the transportation arena. This includes government bodies, transport authorities, public infrastructure operators, insurance, regulated transportation providers, fleet operators and vehicle manufacturers as well as 3rd party contributors. 3rd party contributors, such as those with analytics expertise, would be able to provide insights derived from other data providers, creating new data assets.

As an example, with personal light vehicles starting to incorporate an array of sensor technologies, one could consider the connected vehicle to be a 'sensor platform'. Information gathered by the vehicle is of value, not only to the vehicle manufacturer but potentially to other parties too[2]. A vehicle's sensors are able to determine if it were to drive over a pothole. The location of the pothole could be marked or 'geo-tagged'. The information would be able to guide roadway maintenance teams to conduct the necessary repair, rather than having to use expensive road condition surveys, or reacting to complaints from road users.

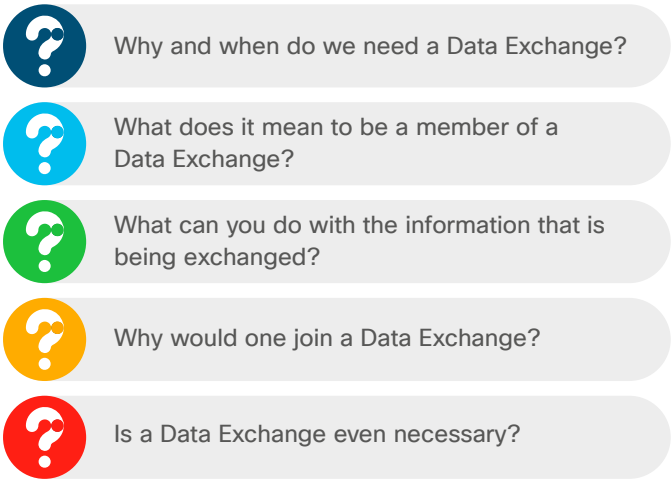
In this document, we'll discuss a framework by which information can be unlocked, enabling the exchange of data between parties in a clear, trusted and verifiable manner. This has the ability to create benefits for citizens, organisations and governments with improvements to safety, sustainability and easing pressing issues such as congestion.

The paper is intended for those responsible for creating data from transportation-related activities or those wishing to leverage data from other providers. This includes public infrastructure operators, private entities, academic researchers, innovators as well as government bodies. The framework is intend to assist in developing strategies and architectures when planning future services. It will review the key pillars required for a successful Data Exchange framework to operate and consider the key role that security plays in such an environment.

The goal is to provide a blueprint for a data framework where information from multiple sources can be exchanged, creating an ecosystem of parties who will be able to build upon and contribute to an exciting future of connected, automated transportation solutions.

[2] https://newsroom.toyota.co.jp/en/corporate/23485508.html

Defining the Data Framework – If it's all connected, why isn't the data?



There are many questions that surround the premise and the operation of a Data Exchange. This section will consider the circumstances under which a Data Exchange may be beneficial.

As transport evolves and changes it is anticipated that the set of participants may well also change over time. Similarly, the use-cases and the data required to support those use-cases will also change. In other words, the change in the nature of a Data Exchange may be organic, triggered by the participants themselves, by the nature and variety of the data that they are able to contribute, in turn generate new use-cases beyond those that may have been in scope originally.

As noted previously, one could design a Data Exchange with a strict set of use-cases and participants in mind. Alternatively, one may look to embrace organic growth and aim to enable as wide a range of current and future use-cases as possible, whilst still providing a framework within which all participants operate.

Looking specifically at multimodal transportation, developments around the world are beginning to demonstrate some of the potential in Data Exchange between entities. For example, when an individual purchases a ticket for a major sporting event, the credit card is tied to a registered location. The seat location within the stadium is also known, therefore it is possible to plan the journey for that customer from their home to their seat and back.

On the match day, the journey plan for that customer can be updated to take into account the latest conditions across road, rail and other forms of transit to provide the optimal route and transport options. At a macro-level, it is possible to take into account other customers travelling from a similar area to provide optimal utilisation of the transport resources. This may result in one set of customers being directed towards particular bus routes, while other fans are directed towards particular train lines and stations.

During the journey, if there are delays on the bus routes due to congestion, customers may be directed to walk towards the nearest tram station in order to get to their seat on time for the start of the match. Through the use of applications on smart devices, customers may be directed towards particular exits at the major stations in order to adapt for optimal passenger flow through the terminus.

The customer experience described above can be achieved if data can be exchanged between the various entities that form the basis for the journey from the customers home to their seat in the stadium.

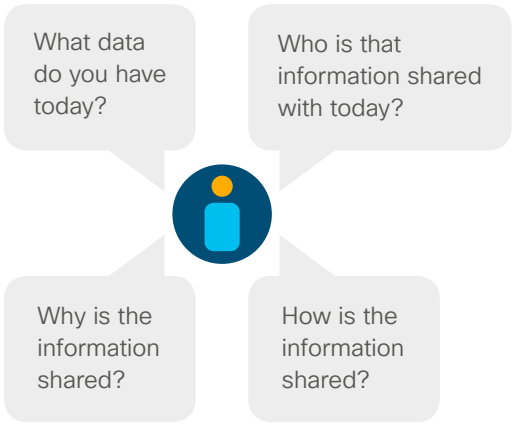
Connected and Automated Vehicles (CAVs) are almost here with many car manufacturers committing to main stream deployment in 2019. This brings with it an immediate need to share data at multiple points within the system. For example, think of the multitude of options when considering V2X. Whether this is within a vehicle, between vehicles, or even from vehicle to infrastructure. Each of these components would perform some sort of Data Exchange function.

Intelligent Congestion Management is another use case that clearly would require benefit from a data framework. In this instance, sensors would determine a congestion event was about to happen yet the impact may be across multiple modes of transport. A Data Exchange could enable the distribution of data to enable the detection of congestion and quite possibly mitigate it through the orchestration of multiple elements within the network.

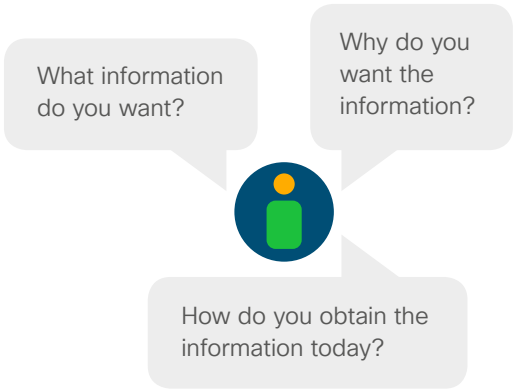
A Framework for Sharing

To enable the sharing of data, there are a number of questions that need to be addressed. Some of these are 'chicken and egg'-like. In order for a consumer entity to come to the exchange, they will want to know what information is available. In order for a producer entity to come to the exchange, they will want to know who the consumers are. The questions of both sides need to be addressed.

To a data producer, one must ask a series of basic questions –



To a data consumer, similarly –



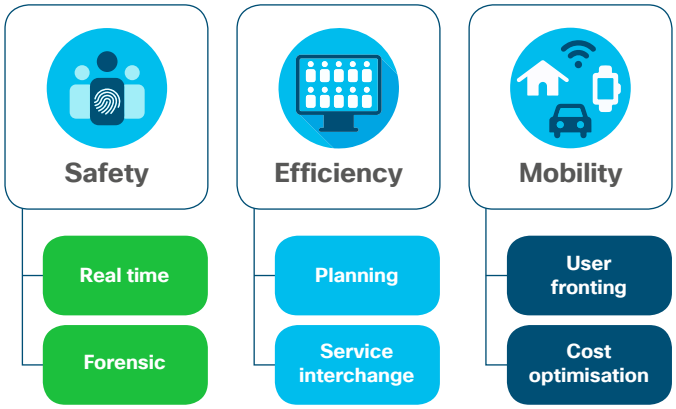
When roadway maintenance authorities plan their work the capability to share that information with public transport authorities and operators enable traffic to be directed away from the affected area rather than waiting for drivers to encounter the roadworks.

What data?

The 'community of interest' will have particular use-cases or outcomes that they are interested in progressing and therefore will have a need for information that contributes towards these goals. At a high level, data in the transportation sector typically falls into one of three categories:

- **Safety** - Data needed for safety purposes or enabling greater safety
- **Efficiency** - Data to assist in planning and services operation
- **Mobility** - Data to assist people, goods and services in moving

Figure 1: Data groupings



The list above is by no means complete and one should expect further categories and sub-categories to be added over time.

WHO?

Next, one needs to understand who the producers and consumers of data are. One must also consider the terms and conditions under which data is being provided, since in some instance, this may be driven by legal frameworks and regulation.

WHY?

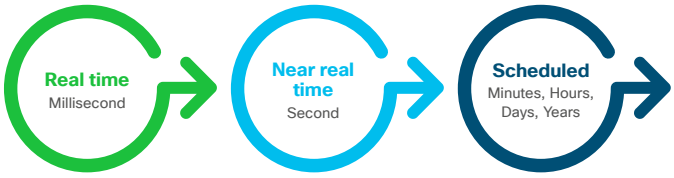
There are many reasons data should be shared. A data owner may be required by regulation to share data with other entities; in some cases it may be through a drive to enable innovation. There may be financial incentives to share data, with various monetisation schemes and mechanisms. Understanding why you are doing something can assist with understanding other aspects of the data sharing, namely, how.

HOW?

Data can fall into 3 loose categories:

1. **Real time data** – generally ephemeral, sub-second (e.g safety messages).
2. **Near real time** – Informational, sub 1-minute (e.g. train updates).
3. **Scheduled** – updates may take 15 minutes (e.g. weather information) or longer.

Figure 2: Data categories



In each case, the data update frequency will likely determine the necessary connector type, transport protocols and distribution relationships. For example, publishing data once per day at a national level, or providing a scoped data-set in near real time to a set of interest parties in a particular geographical region.

Understanding the data

Understanding the data itself will dictate how it should be handled and consumed. In most cases, data is defined by a model that provides structure to the information. In addition, it will typically be presented in a particular format appropriate for the kind of data.

For example, JSON[3] is a commonly used format prevalent in 'web' environments. XML[4] is more structured data format when compared to JSON. However, the format does not provide information about the data itself. For example, train arrival information may be presented, containing information about the train line, train type, train operating company, number of carriages, number of passengers as well as the timestamp on when the train arrived at each station on the line.

Not only does one need to understand what data is being presented but the guarantees, or the service 'contract' that is associated with the data. For example, the train arrival information will always contain information about the arrival at each station on the line (a guarantee) but may not always provide the number of passengers on the train (best effort).

What is the confidentiality of the data?

Data privacy policies dictate what can and cannot be shared, or may define how information can be shared while meeting privacy objectives. This is not the only aspect of data that may cause concern when sharing. Data may expose critical information about governmental operations, provide an unfair competitive advantage in the market place, or reveal data that could be used for nefarious purposes.

Consideration should be given to not only how the data will be used but also how it may be misused. Such an exercise should be performed on a periodic basis, since not only may the set of producers and consumers change but also the nature of the data that is being shared. It is not sufficient to perform the exercise only when the exchange is being planned.

An operating policy should be defined. This might state that parties may share 'enough data to meet the need identified'. This does not mean that additional data cannot be provided in the future but again there should be a verifiable need. Attention must also be paid to the use of identifiable information, be that information pertaining to people or things.

[3] https://www.json.org/

[4] https://www.w3schools.com/xml/

What is the accuracy of the data?

Not all data provisioned is equal in nature. For example, a weather station may provide temperature data rounded to the nearest whole number. Another weather station may provide temperature data to three decimal places. In both cases, the weather stations are providing temperature data but the level of accuracy may be a deciding factor in respect to the data's intrinsic value, depending on the specific use-case. Not only must one consider the accuracy (degree of tolerance) but also the periodicity with which updates are provided and the granularity of temperature samples are taken. A weather station may report the temperature sampled once every five minutes. Another may present an five-minute average derived ten samples, one taken every thirty seconds.

What is the assurance of the data?

What level of trust do we have in the data (and by implication, the data publisher)? A fully secured system may offer good levels of assurance that the data has not been tampered with but the 'chain of assurance' starts with the data's point of origination. Data assurance is primarily a concern of the publisher of the data, since it pertains the checks and balances that the publisher has put in place across the data origination, ingestion and production pipeline. Trust of the infrastructure used to gather the data therefore plays an essential role. If you cannot trust the underlying platforms delivering data, this needs to be reflected in level of trust that you place in the data.

Deriving the metadata

Answers to the questions posed above, form the metadata that needs to be present as part of the Data Exchange environment. Since one cannot envisage all of the potential use-cases that the data will be used for, the exchange needs to provide consumers with the information that will enable them to understand the implication of using that data. Attempting to use data for real-time decision-making when it is only updated on a fifteen minute basis would obviously be hubris. Metadata is therefore essential for consumers to make the appropriate decisions about the use-cases for which data can be used.

For multi-tier informational systems, where data processing is done and the information re-shared, this metadata may change. Greater levels of assurance may be implemented through cross referencing of data. Delays may be introduced. Re-sharing of data must re-evaluate this metadata, preferably adding to assurance and trust levels at each stage.

Who are the community?

Just as all information may not be 'equal' in nature, it is essential to understand who the 'community of interest' is and their motivations, since these will differ, too.

When considering a Data Exchange, it may be relatively easy to identify the parties at the outset since there may be clear common interests in the exchange of data. If there are only a small set of parties that have such a common interest and there is no chance that the scope is likely to grow over time, a 'closed' exchange may be appropriate, meaning an exchange with a clearly defined set of parties and specific use-cases that will be supported. Indeed, if there are only a small set of parties where trust between parties is present, a full Data Exchange facility may not be of value and could be addressed by other solutions such as file-sharing or data 'pipes'.

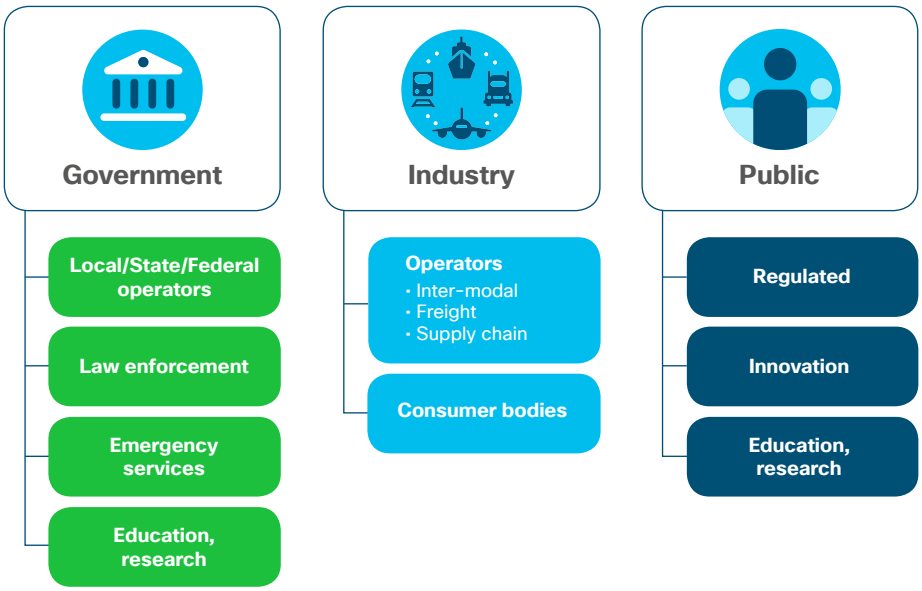
If one looks to examples such as eBay, its success stems from bringing 'sellers' and 'buyers' together for the exchange of value. However, it did not look to limit the set 'sellers' and 'buyers' and rather encouraged a wide and diverse community to join the platform, creating the broad market that we see today.

It is key to note that eBay does operate under strict terms and conditions, applying to both sellers and buyers.

When considering the potential 'community of interest' for the Data Exchange, the salient point from the eBay example is that while there may be a small set of parties whose interests are instrumental in creating a Data Exchange, over the lifetime of the exchange, a successful Data Exchange should expect other parties (both 'sellers' and 'buyers') to appear, attracted by the data and the currently participating entities. In addition, the use-cases that new participants are interested in may be 'different' when compared to the original use-cases that drove the original parties at the founding of the exchange. For example, a data analytics company who've developed algorithms for warehouse automation systems, may apply their expertise in order to develop congestion management algorithms for automated transportation systems.

The diagram below provides a view of potential participants in a transportation Data Exchange:

Figure 3: Potential parties to a Data Exchange



The developer community

There is key constituent that may be overlooked in the establishment of a Data Exchange, namely 'developers' - the group who have to work with the data itself.

When working with data from a single mode of transport, data may well be presented in a specific format that supports the expected use-cases and consumption models. When considering a multimodal situation, it is very likely that a wide variety of data models and formats will be in use, creating a challenge for the development community since they will need to be able to consume data from the variety of different source and modify the format in order to suit their needs.

In the best case, this may be quite a simple task due to a very limited set of data formats and models. However, if we think towards an evolving exchange, the challenge will become the on-boarding of additional participants and the need to then incorporate the new data that's being offered. For example, an exchange may start by offering Bus schedules and 'real-time' arrival data from a range of Bus operating companies. At a later point, 'on-demand' pedal bike information could be added, including 'real-time' bike location and availability data.

Complexity is therefore very likely to increase over time as parties look to build 'adapters' to work with the range of data types that are on offer. This may dissuade potential participants as there will be an ongoing maintenance cost associated with each of the adapters that the participants will need to create.

While there are standards such as 'The Service Interface for Real Time Information' (SIRI)[5] and 'General Transit Feed Specification' (GTFS)[6], one should not expect that participants are already producing data to these specifications, not that these specification will be the only ones that will be used over the lifecycle of the exchange.

If we look to the example of the adoption of Application Programming Interfaces (APIs), successful APIs typically have a number of key targets that they look to deliver for their development community:

These tenants are key when considering how a Data Exchange will function.

[5] https://www.vdv.de/siri-white-paper08.zipx?forced=true

[6] https://en.wikipedia.org/wiki/General_Transit_Feed_Specification#cite_note-1

Goals and principles for a Data Exchange

The primary purpose of a Data Exchange is just that, the exchange of data. The following proposed goals that should underpin a Data Exchange:

- Availability of high quality data with consistency of service - publishers adhere to guarantees for availability of data that consumers can rely upon
- Underpinned by the required legal and license structures - respect for the concerns of data publishers and applicable legislation
- Utilisation of industry-aligned, standards-based interfaces - maximise interaction and interoperability

In the creation of the technical exchange infrastructure and functionality, the following principles should drive the design and operation:

- Enable maximal use and application of the data
- Ensure access with control
- Present data in its most interoperable and most computable formats
- Offer a unified and consistent developer experience over disparate, heterogeneous datasets

Premise of operation

When considering the concept of a Data Exchange, it is helpful to look at the model of Internet Exchange Points (IXPs). IXPs (typically) represent a physical location where (commercial) entities are able to come together on 'neutral' territory. Entities may choose to connect to 'common data infrastructure', owned and operated by the IXP, enabling communication between entities on a many-to-many basis. Entities may also choose to connect to selected entities on a 'point-to-point' basis. In other words, private communication that happens to take place within a 'neutral' facility.

Entities who choose to use IXPs pay to be a member of the organisation and pay for power, space and cooling for any equipment that they choose to place within the physical premises. Members will have the ability to participate in the governing body overseeing the operation (commercial and technical) of the IXP.

The 'value' of the exchange comes from the volume and variety of its membership as well as the cost reduction of working through an exchange versus establishing and maintaining point-to-point connections between entities with a valuable commercial relationship.

Further, other industrial sectors have identified the need for Data Exchanges, establishing the exchange of specific types of information or for specific kinds of entities, for example, pharmaceutical information being exchanged between manufacturers and medical authorities. Further, the potential to establish Data Exchanges at a country or regional level.

There are already examples of Transportation Data Exchanges in operation today. However, in most cases the exchange is owned and operated by a Public / Government Transportation authority. Membership of the exchange is predicated on the entities being contracted to operate services for the Transportation authority. Data is exchanged between transportation service operators (for example, a Bus operating company) and the Public Transportation authority as part of the framework under which the operating contract is being executed.

In such scenarios, the Public Transportation authority is both the operator of the exchange and is the primary user of the data. They too may need to publish data to other branches of Government or to the general public as part of established policies or requirements. It is however, unusual to see data being exchanged between transportation service operators (Bus operating company to Rail operating company or between Bus operating companies) through such a facility.

Commercial entities are able to leverage the publicly-published data from the Public Transportation authority but do so outside of the formal Data Exchange.

Table 1: Comparison of Internet Exchange Point and Transportation Authority Data Exchange

	Internet Exchange Point	Public Transportation authority
Owned and managed by entity	Yes	Yes
Membership on open (commercial) basis	Yes	No
Members have voting rights	Yes	No
Member to Member Data Exchange	Yes	No
Exchange operated for profit	In some cases	No
Policies govern use of exchange	Yes	Yes
Policies define data to be exchanged	No	Yes

The table above provides a high-level comparison of the operating principles under which these two forms of exchange operate. It must be noted that nothing prevents entity to entity Data Exchange taking place outside of the auspices of either exchange model.

A Transportation Data Exchange, operated along the lines of an Internet Exchange Point would, by the nature of the construct, enable different Data Exchange relationships. With membership on an open basis, it would be possible for both public and private/commercial entities to 'meet' at such a location. This would enable Data Exchange between entities such as freight haulage companies, public roadway authorities and rail service providers. In such a scenario, one could see freight vehicle information being used in conjunction with roadway congestion state information and rail freight-terminus loading/unloading schedules to create improved streamlined 'last-mile / first-mile' services as well as improvements in fuel efficiency and reduced environmental impacts.

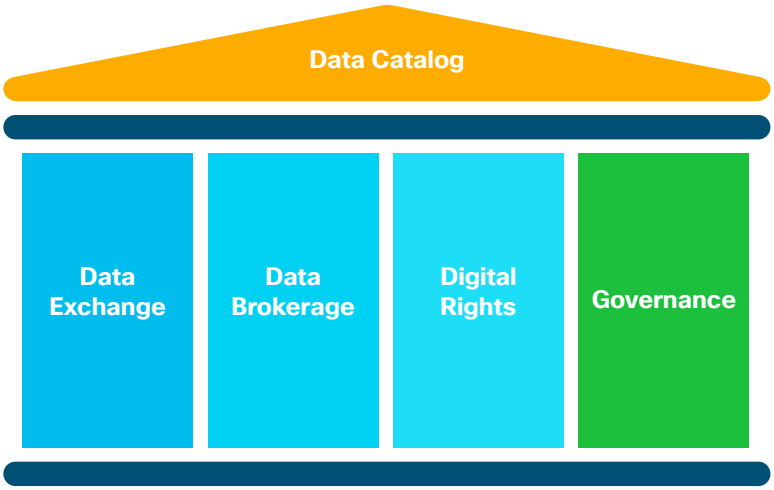
Furthermore, with the appropriate commercial structures in place, it would be possible to encourage 3rd party entities such as data analytics companies to become part of the value chain, providing their expertise in creating new data assets that would be value to other members of the exchange. For example, optimised freight loading for freight rail-cars based on detailed information of the loads being transported by the freight haulers.

It is this opportunity for improved multimodal transportation coordination that such an exchange would be best positioned to deliver.

Laying the Foundations of the Data Framework

The diagram below contains the foundations that support a successful Data Exchange. Each of these elements serves a specific function that will be described in more detail:

Figure 4: Essential pillars of a Data Exchange



Data Exchange

The exchange represents the technology through which data from a publisher can be obtained by a consumer. In its simplest form, this may be a function provided between a pair of entities but with the goal of being able to support an ecosystems of publishers and consumers, the technology needs to enable transactions between many parties and be able to support the exchange of a wide variety of data.

Security of Data Exchange as well as the authorisation and validation to participate is a fundamental operational function of the exchange.

A Data Exchange is intended to bring multiple parties together for the exchange of data. One design approach is to require all publishers to contribute their information into a centralised data-store or set of data-stores that reside outside of the entities domain. Consumers are then able to access the data-store, subject to policy and permissions.

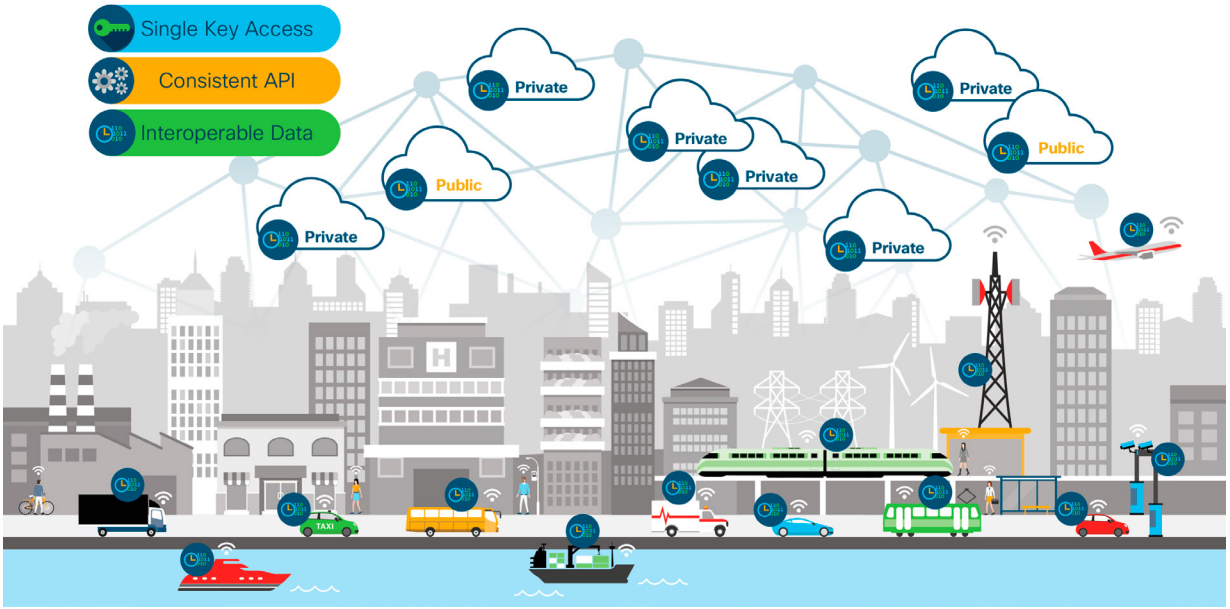
An alternative design approach is to enable publishers to connect to an exchange, with data remaining within the realm of control of the publisher, being retrieved on demand when a consumer requests information. In this method, data remains distributed.

There are of course, advantages and disadvantages of each design approach. Creating a new data-store provides an opportunity to impose a format or standard to the data that is being contributed. However, it also creates the need for lifecycle management of the data-store. With the distributed approach, the publisher retains ownership and responsibility for the data up to the point where the information is provided to the consumer. Given the range and variety of data sources, it is very likely that the data will be present in multiple different formats, creating an operational burden for the consumer.

There are many advantages and disadvantages beyond those outlined above. A key point to consider is the 'cost of entry'. If a potential data publisher must convert their information into another format in order to participate, it becomes a burden to both build and maintain that capability. For a consumer, having to learn to address the variety of data formats creates a barrier, especially if every time a new data source is added, the consumer has to adapt accordingly.

The issues of both publisher and consumer can be addressed.

Figure 5: Distributed system, unified access design



The design above proposed a 'distributed system' approach. Publishers make their existing data available in its original format. Consumers are presented with a single entry point to the Data Exchange and within it, have access to a single, well-structured API. The 'middleware' bringing these two sides together are a series of connect and transform functions that translate data from its original format to a small set of standardised format, such as JSON or MQTT.

In this model, data continues to reside with the publisher. A consumer will make a request for data, that request being converted into the necessary query logic that the publisher's systems can understand. The returned data will be converted by the connect and transform function from its original form to the standardised form.

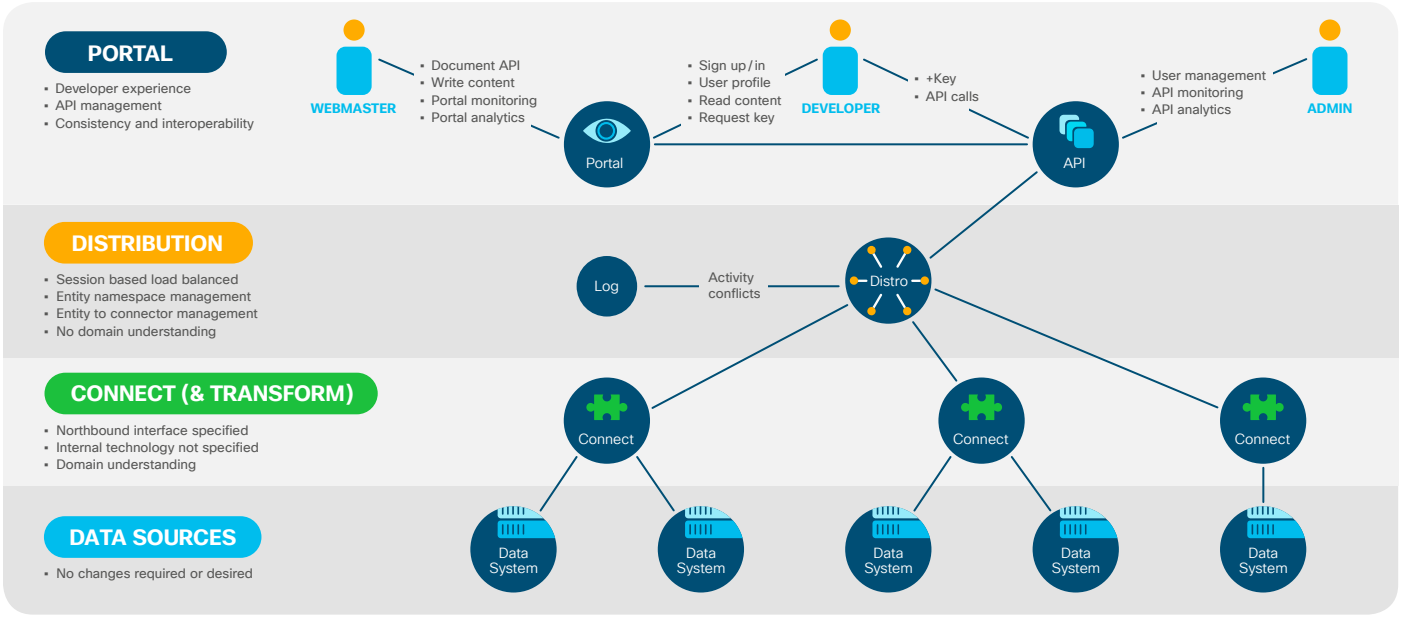
A fundamental requirement is for the data publishers to be able to acquire information in the first instance. Expecting entities to make large initial investments in the necessary data acquisition infrastructure purely for the purposes of participating in the Data Exchange is unrealistic. Publishers should be encouraged to share data that they already retain. If they do not have data acquisition capabilities in place, an appropriate analysis of their internal use-cases and business needs should be undertaken before considering connection to the Data Exchange.

In practical terms, data publishers will have data obtained from a range of sources such as in-vehicle sensors, roadway and track condition monitoring systems and WiFi/cellular analytics solutions. The necessary data acquisition infrastructure will be required to obtain and process the data in support of an entities business goals.

Participation in the Data Exchange should bring additional value to the publishing organisation. Economic benefit may encourage the publisher to increase the volume and variety of data being acquired and in turn, the data being published via the Data Exchange.

Another key component to realising the Data Exchange is the logic to be able to determine which publishers to query for data, considering that for example, personal light transport data would be provided by multiple publishers. The knowledge of which publishers can contribute to each type of data resides within a function known as the distribution layer. Its purpose is to translate a query from the consumer into queries that will be directed to the appropriate publishers.

Figure 6: Functional layers of the Data Exchange



Above the distribution layer are the elements that the consumers will engage with in the process of obtaining access to data. As noted previously, it has been shown that for a platform to be successful, there must be an intense focus on the needs and wants of not just the consumers but the development organisations within the consumer entity.

The portal layer provides three key functions; the developer portal, the API portal and the API management element. These provide the ability to monitor and manage the operation of the Data Exchange, the users of the exchange and the data to which they are able to get access to. In addition, it provides the overall platform management functions including the operation of the connect and transform layer.

Securing the Data Exchange

Whether the Data Exchange is based on a centralised data store or a distributed one, the Data Exchange must be protected against threats. The security of the data begins with the protection of the data sources themselves and the devices used to enable the exchange of data. With the appropriate protections of these sources, the devices and applications can become reliable producers and consumers in the Data Exchange.

As described in Figure 6, with the secured data sources and devices, the connectivity and transform layer must ensure that the communications are also protected using validated cryptographic tools to provide the appropriate levels of confidentiality and message integrity. Data transformations may also be required (prior to encryption) to further ensure privacy. Since all communications are performed between connectivity and transform layer and the individual data sources, rather than between data sources, the information is unicast. This means that each party can establish secure communications directly. The actual means for affecting the communications security may vary. For instance, if symmetric key cryptography is used, the Data Exchanged secured by that key only affords the knowledge that someone in the group may have been the publisher; while asymmetric cryptography can be considered, its computational costs make it a less palatable option.

As is the case with most internet-based services today, a registration or enrolment will be required to ensure that only authorised devices and applications can act as publishers, consumers or both in the Data Exchange. To automate the Data Exchange, and to affect strong communications security, the devices and applications should be provisioned with strong credentials (preferably X.509 certificates) to enable them to establish themselves as authorised participants in the Data Exchange.

Data Catalog

As noted previously, one of the challenges with establishing a Data Exchange is the need to bring both consumers and producers together. From potential consumers, it is typical to hear the statement 'I don't know what information is available' or 'I never knew that anyone published this data'. Likewise, from potential producers, it is not unusual to hear 'No one would use this information anyway'.

A data catalog is an essential component of a Data Exchange, providing parties with the ability to discover what information is available. In addition, the metadata is an essential element, enabling consumers to be able to see the terms and conditions associated with the use of data and the service guarantees that the producer is abiding by. It does not make sense for a consumer to attempt to build a service aiming to offer real-time information if the underlying data source is offered as a 'best-effort' data stream. If data is being offered on a 'pay-per-update' basis, that needs to be stated.

The catalog must be a dynamic reflection of the data that the exchange offers, showing the breadth of information available. For example, passenger numbers by bus route, train line, station, by time of day, weather condition information, road surface conditions, parking bay availability, Electric Vehicle charging point location and utilisation etc.

Data Brokerage

Data is increasingly being seen as of value, and in some cases, value that can be realised through monetary exchange. An entity may establish a contract with a data provider for a certain set of data for a period of time with the entity paying for that data. Establishing contracts with the range of publishers and consumers can be a considerable challenge, and indeed, not all data may be considered to be of monetary value. Further, a publisher may choose to charge one consumer for access but may make the same data available to another entity for free.

In some cases, publishers may recognise that there may be value in the data that they are offering but are not able to price the data.

A brokerage entity, a neutral third-party, can act as a central clearing point. It may take on the duties of confirming the validity of the exchange of data between parties as well as establishing a marketplace for trading data. The financial model that would support the brokerage entity is beyond the scope of this paper.

Digital Rights

Policies describing the intent of the Data Exchange as well as supported uses also form valuable documents when considering if use-cases are potentially straying outside the realm that the exchange is intended to enable.

As noted previously, the terms and conditions associated with data that a publisher is making available are fundamental to its use. At a deeper level, attribution and terms of ownership of data must be clearly stated up front. The rights associated with data being provided by publisher to consumer, who in turn produces derived works from that original data must be clearly established.

Traceability of information is area that should be considered, too. It may be required by regulation to provide provenance of data as well as a chain of custody that can be examined.

Care and attention must also be paid to data that may contain 'Personally Identifiable Information'[7]. While each publisher must be responsible for adhering to regulation and legislation pertaining to data, a clear policy should be established in respect to the data being exchanged through the Exchange's auspices. It is not proposed that the Exchange should examine the data that each publisher is making available.

Governance

As noted previously, models such as Internet Exchange Points typically have a group responsible for the organisation and operation of the exchange. Such a group is responsible for establishing and enforcing policy over the operations conducted by the exchange. A similar function is proposed for a Data Exchange.

[7] https://en.wikipedia.org/wiki/Personally_identifiable_information

Conclusion and Recommendations

Transport systems are in the early stages of embracing digitisation. While it is perfectly reasonable for individual transport mode operators to be exclusively focused on the benefits that digitisation will bring to their domain, there is significant value that will be missed if the opportunities that transportation Data Exchanges bring are not realised.

Government bodies should be considering the frameworks and policies necessary to facilitate the secure, controlled exchange of transportation data between public and private entities. Those frameworks and policies should look to encourage rapid innovation, prototyping and entrepreneurism in order to attract the broadest range of participants.

It is strongly advised that a regional approach be adopted as a starting point, with strong inter-regional coordination. Regional Data Exchanges should then form under a national (and potential international) Data Exchange.

Attention must be paid to the risk of bifurcation, leading to incompatibilities between exchanges.

A strong effort will be required to bring in the initial set of data contributors. In parallel, efforts will be required to build and maintain a vibrant development community who are active in building value from the data present within the exchange.

Those countries or organisations who are prepared to build Data Exchanges and contribute data will have significant “first mover” advantage and will be able to recognise accelerated economic Transportation benefits.

Cisco woud like to acknowledge the contribution of the following people:

Jeff Sharp (Downer); Majid Sarvi (University of Melbourne); Fred Wang (Public Transport Victoria); Ross Caldow (Here Maps); Mark Lawler (Transurban); Vettri Ganesan (Transurban); Shane Hazim (Australia Post); Craig Sampson (Australia Post); Gary Desillan (Australia Post); Jamie Smith (Telstra); Wayne Glenn (WSP); John Tran (Brisbane City Council); Tanya Pennell (Transport Main Roads); Igor Santic (Transport Main Roads) Matt Kluck (Transmax); Thomas Yeo (TfNSW); Niro Jeyarajah (TfNSW); Meead Saberi (UNSW); Andrea Ash (Cohda Wireless); Jay Cheah (Cohda Wireless); Jonathan Myers (Keolis-Downer); Niko Limans (Austroads); John Luan (Intellimatics); Russell Brett (IAG); Matt Howe (Infranexus); Ben Itzstein (Data61); Cory Ross (Main Roads)